

Genome Analysis of a *Glossina pallidipes* Salivary Gland Hypertrophy Virus Reveals a Novel, Large, Double-Stranded Circular DNA Virus[∇]

Adly M. M. Abd-Alla,^{1,3*} François Cousserans,² Andrew G. Parker,¹ Johannes A. Jehle,⁴
Nicolas J. Parker,⁵ Just M. Vlak,⁶ Alan S. Robinson,¹ and Max Bergoin²

Entomology Unit, FAO/IAEA Agriculture and Biotechnology Laboratory, Agency's Laboratories, A-2444 Seibersdorf, Austria¹;
Laboratoire de Pathologie Comparée, Université Montpellier II, Montpellier, France²; Department of Pests and Plant Protection,
National Research Centre, Dokki, Giza, Egypt³; Agricultural Service Center Palatinate (DLR Rheinpfalz), Breitenweg 71,
67435 Neustadt a/d Weinstrasse, Germany⁴; 10 Lockhart Close, Kenilworth, Warwickshire CV8 1RB,
United Kingdom⁵; and Laboratory of Virology, Wageningen University, Binnenhaven 11,
6709 PD Wageningen, The Netherlands⁶

Received 5 December 2007/Accepted 1 February 2008

Several species of tsetse flies can be infected by the *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV). Infection causes salivary gland hypertrophy and also significantly reduces the fecundity of the infected flies. To better understand the molecular basis underlying the pathogenesis of this unusual virus, we sequenced and analyzed its genome. The GpSGHV genome is a double-stranded circular DNA molecule of 190,032 bp containing 160 nonoverlapping open reading frames (ORFs), which are distributed equally on both strands with a gene density of one per 1.2 kb. It has a high A+T content of 72%. About 3% of the GpSGHV genome is composed of 15 sequence repeats, distributed throughout the genome. Although sharing the same morphological features (enveloped rod-shaped nucleocapsid) as baculoviruses, nudiviruses, and nimaviruses, analysis of its genome revealed that GpSGHV differs significantly from these viruses at the level of its genes. Sequence comparisons indicated that only 23% of GpSGHV genes displayed moderate homologies to genes from other invertebrate viruses, principally baculoviruses and entomopoxviruses. Most strikingly, the GpSGHV genome encodes homologues to the four baculoviral per os infectivity factors (p74 [pif-0], pif-1, pif-2, and pif-3). The DNA polymerase encoded by GpSGHV is of type B and appears to be phylogenetically distant from all DNA polymerases encoded by large double-stranded DNA viruses. The majority of the remaining ORFs could not be assigned by sequence comparison. Furthermore, no homologues to DNA-dependent RNA polymerase subunits were detected. Taken together, these data indicate that GpSGHV is the prototype member of a novel group of insect viruses.

Tsetse flies (*Diptera: Glossinidae*) are responsible for the transmission of two major zoonoses in tropical Africa: human African trypanosomiasis (sleeping sickness), which is currently estimated to infect 300,000 to 500,000 people with another 60 million at risk in 36 countries, and animal African trypanosomiasis (nagana), which causes heavy losses of cattle, domestic pigs, and other farm animals (69, 81). Both diseases have profound impacts on agricultural development and nutritional resources in sub-Saharan Africa (37). Due to the lack of effective vaccines and efficient, inexpensive drugs for the human disease and the development of drug resistance in the animal disease (3), the control of the vector populations remains the main strategy for controlling these scourges.

The successful eradication of *Glossina austeni* from the island of Unguja, United Republic of Tanzania, using an integrated approach including the release of sterile male tsetse flies (77) stimulated the development of programs aimed at applying this approach on the African mainland, such as one to control *Glossina pallidipes* in the Southern Rift Valley in Ethiopia (22). The establishment of a laboratory colony of the

target species and then its mass production in large-scale facilities are sine qua non prerequisites for the successful implementation of the sterile insect technique. In the framework of the sterile insect technique program in Ethiopia, pupae of *G. pallidipes* from the target area were shipped to the Entomology Unit of the FAO/IAEA Agriculture and Biotechnology Laboratory, Seibersdorf, Austria. A colony was established, which reached 15,000 females by July 2000 but experienced a steady decline in 2001 and was lost by 2002. Efforts to identify the cause(s) of the decline led to the discovery of a high proportion (over 85% in some samples) of male and female individuals with salivary gland hypertrophy (SGH), a syndrome first described in wild populations of *G. pallidipes* (12, 80) but later observed in many tsetse fly species from different African countries (17, 24, 38, 39, 48, 50, 54, 65). The causative agent of this syndrome was later identified as a nuclear rod-shaped, enveloped DNA virus averaging 70 by 640 nm in size (33). This virus was also found to be associated with testicular degeneration and ovarian abnormalities (38, 42, 59, 60), and its presence has been shown to affect the development, survival, fertility, and fecundity of naturally (20) or experimentally (40, 62) infected flies. Mother-to-offspring transmission, either transovum or through infected milk glands (39, 59, 61), is thought to be the mode of transmission of the virus in natural tsetse fly populations. In reared colonies, however, horizontal transmission facilitated by the membrane feeding technique used for

* Corresponding author. Mailing address: Entomology unit, FAO/IAEA Agriculture & Biotechnology Laboratory, IAEA Laboratories, A-2444 Seibersdorf, Austria. Phone: 43 1 2600 28425. Fax: 43 1 2600 28447. E-mail: a.m.m.abd-alla@iaea.org.

[∇] Published ahead of print on 13 February 2008.

large-scale feeding of tsetse flies (21) may constitute the main route of virus transmission. The SGH syndrome was also observed in two other species of nonhematophagous diptera, the narcissus bulb fly *Merodon equestris* (6) and the housefly *Musca domestica* (15).

The correlation between the high incidence of SGH syndrome of the Ethiopian *G. pallidipes* colony in the Seibersdorf laboratory and its collapse prompted research to develop a sensitive assay to unequivocally identify the flies carrying the virus. Based on previous work with partial characterization of purified virions isolated from *G. pallidipes* showing SGH (51), a highly sensitive, nondestructive diagnostic PCR method was developed that revealed a very high (up to 100%) prevalence of the virus in asymptomatic flies (1). As a first step to obtain a better understanding of the virus-host interaction and eventually to develop ways to manage the virus in tsetse fly colonies, we isolated the virus and sequenced its complete genome. We report here the entire nucleotide sequence and annotation of the GpSGHV genome. Although sharing some properties with other large, double-stranded, circular DNA viruses of invertebrates, analysis of the 190,032-bp large genome of GpSGHV revealed that the virus cannot be assigned to any of the so far described families of DNA viruses and thus should be considered the prototype of a new type of insect virus family.

MATERIALS AND METHODS

Isolation, cloning, and sequencing of GpSGHV genomic DNA. Virions of GpSGHV were purified by sucrose density gradient centrifugation from hypertrophied salivary glands collected from infected *G. pallidipes* flies that were maintained in the FAO/IAEA Agriculture & Biotechnology Laboratory, Seibersdorf, Austria (1). Intact genomic DNA was extracted from purified virions, and a partial DNA library was produced by cloning EcoRI fragments in pUC19 plasmids using *Escherichia coli* TG1 bacteria (1). The lengths of cloned inserts were estimated by EcoRI digestions followed by electrophoresis in a 1% agarose gel. A total of 415 recombinant plasmids with inserts ranging from approximately 1 to 7 kb were selected. DNA templates were sequenced from both ends with M13 universal sense and antisense primers (MWG-Biotech AG, Germany) using the Sanger method (63). Moreover, 3 μ g of intact DNA was sent to 454 Life Sciences (Bradford, CT) for pyrophosphate-based sequencing (45). A total of 321,953 singleton reads and 402 contigs were provided by 454 Life Sciences, generated by their own assembly software (Newbler). Most of these reads (90%) averaged 80 to 139 nucleotides (nt) in length, summing to approximately 160 times the estimated length of the GpSGHV genome. Assembly of these contigs and the sequences obtained from the cloned EcoRI fragments was performed using Phrap, SeqMan (Lasergene 7.0; Dnastar Inc.), and NTI Vector 9 (Invitrogen) software packages. Additional contigs were assembled from the individual reads provided by 454 Life Sciences using simple custom-made tools (54a). Gaps and ambiguous regions were amplified by PCR using appropriate primers and viral DNA as the template and sequenced. The combination of shotgun pyrosequencing, the Sanger sequencing of EcoRI restriction fragments, and PCR gap closure reactions generated a circular DNA consensus sequence representing on average a 10-fold redundancy at each base position.

DNA sequence analysis. DNA sequence data were analyzed using the GeneQuest (Lasergene 7.0; Dnastar Inc.), NTI Vector 9 (Invitrogen), and ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) programs. Open reading frames (ORFs) corresponding to more than 50 amino acids with a methionine start codon and with minimal overlap (<100 nt) were evaluated for their putative protein-coding potential. All the predicted ORFs were analyzed using BLAST programs at the NCBI website (<http://www.ncbi.nlm.nih.gov>) against the nonredundant protein database (5), the Viral Bioinformatics Resource Center (<http://athena.bioc.unvic.ca/>) database, and a viral protein database that we established by downloading the *Ascoviridae*, *Asfaviridae*, *Baculoviridae*, *Herpesviridae*, *Iridoviridae*, *Nimaviridae*, *Nudivirus*, *Polydnaviridae*, and *Poxviridae* protein sequences of GenBank's nonredundant protein database. Similarities to viral or cellular genes were accepted on the basis of BLASTP (default parameters) scores of ≥ 44 and ≥ 50 , respectively.

Protein motifs were analyzed by using the PROSITE database, release 16 (31).

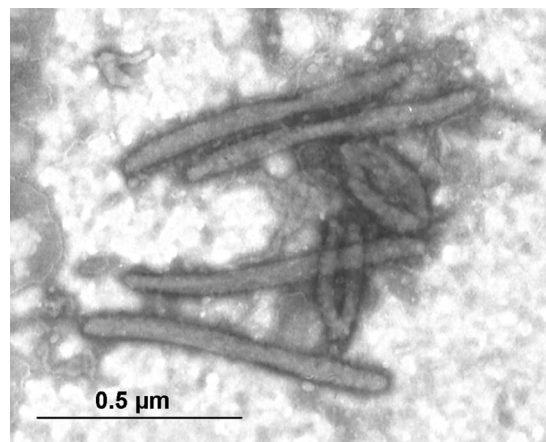


FIG. 1. Virions of GpSGHV observed by electron microscopy after negative staining.

The sequence was analyzed for repeats using GeneQuest and Tandem Repeats finder (9). Multiple alignments were performed using BioEdit and ClustalW, and the percent amino acid identity indicated the percentage of identical residues between complete ORFs. Promoter analysis (100 and 200 nt) for motif searches was carried out using a set of computer scripts made in the computer scripting language Perl. In order to establish the relationship of GpSGHV to the large double-stranded DNA (dsDNA) virus families, phylogenetic trees were constructed using ClustalW protein alignment and MEGA 4 (70) software with neighbor joining and were confirmed by bootstrap analysis with heuristic search and 500 replicates.

Nucleotide sequence accession number. The complete nucleotide sequence of GpSGHV was deposited in GenBank under accession no. EF568108.

RESULTS AND DISCUSSION

General features of the GpSGHV genome. The viral DNA was prepared from purified virus particles (Fig. 1) isolated from a pool of heavily infected salivary glands as previously described (1). The sequencing of GpSGHV DNA by a combination of (i) shotgun pyrosequencing, (ii) Sanger sequencing of a partial genomic library of cloned EcoRI fragments, and (iii) sequence gap filling using PCR products, followed by sequence assembly, led us to conclude that the GpSGHV genome consists of a circular dsDNA molecule of 190,032 bp. The size is in good agreement with that of 185 to 220 kbp estimated by summing the sizes of the fragments generated by EcoRI-restricted viral DNA or measured by pulsed-field gel electrophoresis of intact DNA molecules (1). Furthermore, the sizes of the BglIII restriction fragments measured from an agarose gel (Fig. 2) fit well with those predicted from the genomic sequence and no submolar bands were observed, indicating that the viral DNA was pure and that sequencing and assembly were carried out accurately. As mentioned above, the DNA used for pyrosequencing was not clonal but was extracted from a large virus population, and thus it is very likely heterogeneous. Indeed, the very large number of reads (321,953) provided by pyrophosphate sequencing allowed us to detect 101 positions with nucleotide polymorphism distributed along the whole sequence (Table 1). The most frequent nucleotide was taken into account to establish the viral sequence. The GpSGHV genome is one of the largest known genomes of insect viruses with circular dsDNA genomes, exceeded only by that of *Heliothis zea* nudivirus 1 (228 kbp) (14).

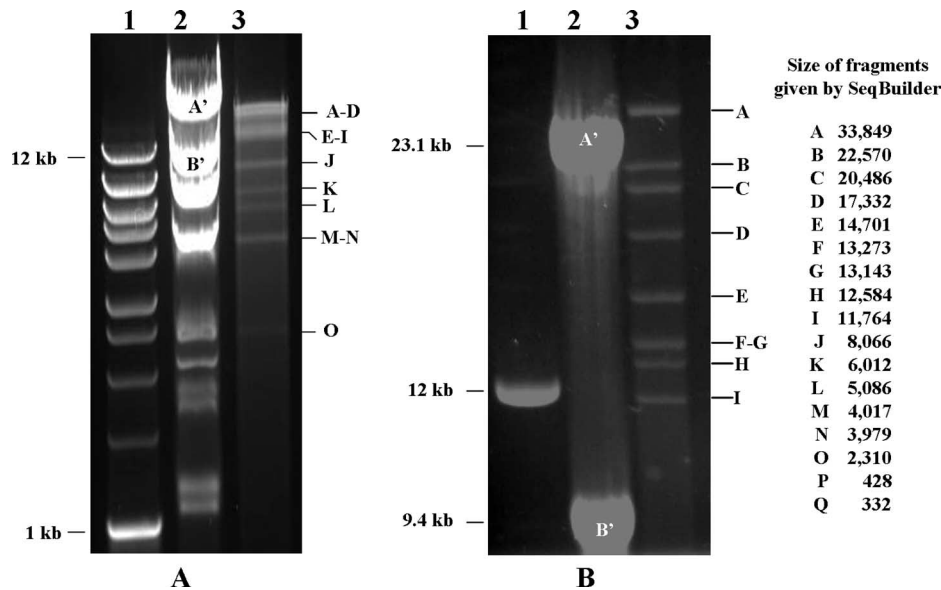


FIG. 2. Electrophoretic profile of BglIII-restricted GpSGHV DNA, showing 15 of the 17 restriction fragments generated by this enzyme. (A and B) Agarose gels after 3 h and 24 h of migration, respectively. The two small fragments P and Q (<1 kb) cannot be seen on the gels. Submolar bands were not apparent but may be present at a lower concentration. Lanes 1, 1-kb DNA ladder; lanes 2, Hind III-restricted λ DNA ladder (A' and B) correspond to 23,130 and 9,416 bp, respectively); lanes 3, GpSGHV DNA.

The very low G+C content of 28% of the GpSGHV genome significantly differs from those reported for most invertebrate viruses with large circular dsDNA genomes, including baculoviruses (33 to 58% G+C) (36), nudiviruses (42% G+C), (14, 79) (with the exception of the cricket nudivirus GbNV [28% G+C] [79]), nimaviruses (41% G+C) (75, 84), and ascoviruses (35.5 to 49% G+C) (7, 10, 78). It also differs from those of the large linear dsDNA genomes of herpesviruses (35 to 75% G+C) (74) and of most iridoviruses (49% G+C) (73), except *Chilo* iridescent virus and lymphocystis disease virus (both 29% G+C) (34, 72), but is closer to the low (18 to 27%) G+C content of entomopoxviruses (2, 8).

Gene content analysis. A total of 322 potential ORFs with a methionine start codon and corresponding to a minimum protein size of 50 amino acids were detected in the GpSGHV genome. Among these, 160 ORFs with no or minimal overlap (≤82 nt) were assumed to encode putative proteins (Table 2). The predicted ORFs are evenly distributed on both strands (51% forward and 49% reverse); many of them are arranged in unidirectional gene clusters (Fig. 3). They represent 86% of the genome, with an average density of one gene per 1.2 kb. The A of the ATG start codon of ORF SGHV001, encoding a homologue of a baculovirus p74 occlusion-derived virion (*odv*) envelope protein, was arbitrarily assigned as position 1 of the GpSGHV genome sequence.

BLAST searches revealed that 47 of the 160 predicted GpSGHV ORFs had putative homologues in the NCBI protein database. These include 28 putative homologues to putative or known genes from other large dsDNA viruses, essentially baculoviruses (12 ORFs), entomopoxviruses (16 ORFs), and 10 ORFs homologous to putative or known genes from cellular organisms (Table 2). The level of amino acid identity between the SGHV putative ORFs and those of other viruses was usually low, ranging from 13.8% to 37.4% with an average

of around 22%. The remaining 113 predicted ORFs produced no significant BLASTP hits.

We also studied the relative abundance of putative regulatory elements in the promoters of the ORFs by computational analysis of the SGHV genome. This enumeration analysis was carried out using the program Perl for all 4- and 5-nt motifs in the 100- and 200-nt upstream sequences (Table 3). The method was calibrated on extensively studied large DNA viruses such as herpesviruses, baculoviruses, and vaccinia virus (46). Analysis of SGHV revealed that the 4-nt motif TAAG has the highest relative enrichment (3 times; $P \leq 0.05$) in the 100-nt promoter compared to the whole SGHV genome (Table 3). For the 5-nt motif, ATAAG and TAAGA were highly enriched (4.9 and 3.7 times, respectively; $P \leq 0.05$). This suggests that the TAAG motif may have a regulatory function in transcription. The same motif is part of the major late transcription initiation site of baculoviruses, and this may suggest an ancient regulatory relationship between GpSGHV and baculoviruses. However, no transcriptional analysis has been done to support the hypothesis that TAAG plays a role in SGHV transcription.

Repetitive regions. There are 14 repetitive regions (RS) with head-to-tail tandem repeat sequences (TRS) and one inverted repeat sequence distributed throughout the GpSGHV genome, representing about 3% of the genome sequence (Fig. 4). The RS are 171 to 556 nt in length and consist of 78 TRS minifragments. Most of these TRS are highly homologous to each other and are clustered in two genome areas located in the 86- to 88-kb and 179- to 183-kb regions. The sizes of the TRS varied from 52 to 246 nt, and the number of TRS per RS varied from 2.7 to 14.5. Within the same RS, the identity of TRS was over 80%, but the sequence identity among the different RS varied from 21.2 to 96.2%. TRS 9, 11, 13, and 15

TABLE 1. Sequence polymorphisms in the GpSGHV genome

Position	Variation		Position	Variation	
	Nucleotides	Abundance		Nucleotides	Abundance
13312	A/T	53/4	82549	G/A	44/1
13322	C/T	53/4	82941	G/A	62/1
13337	C/T	53/4	83042	A/G	58/3
13346	G/T	53/4	83062	T/A	58/3
14631	C/T	54/4	83064	T/G	58/3
16740	G/A	18/2	83066	G/A	58/3
16870	A/G	29/2	83069	A/G	58/3
16894	A/G	28/3	83071	A/G	58/3
20645	T/A	47/3	83073	A/T	58/3
20692	A/G	48/2	83125–83126	CG/TA	41/1
22920	C/T	45/4	83137	T/C	41/1
23169	C/T	40/1	83158	C/T	36/2
53914	C/G	23/3	83204	G/A	59/1
66279	A/G	61/1	83223	C/T	58/2
66341	T/A	55/2	83237	A/G	59/1
66449	C/T	69/3	83277	G/A	43/3
81140	G/A	62/17	83289	G/A	52/2
81362	A/C	81/1	83302	A/G	52/2
81411	T/C	48/2	83411	A/C	61/4
81413	A/G	48/2	83448	G/A	53/2
81427	T/A	48/2	83644	T/C	78/2
81434–81435	GT/AC	48/2	83665	T/C	78/2
81440	A/C	82/5	83674	A/G	78/2
81456	C/A	83/4	83683	A/G	78/2
81470	G/A	84/3	83711	T/C	49/2
81480	G/C	84/3	86544	T/C	69/4
81510	G/C	69/2	86577	T/C	69/4
81514	G/T	68/3	86582	A/G	69/4
81532	G/A	69/2	86584–86585	AT/CA	69/4
81556	A/G	80/7	86592	A/G	61/2
81561	A/C	80/7	86594	A/G	61/2
81572	A/T	80/7	86596	T/C	61/2
81638	G/A	37/3	90097	A/G	110/4
81664–81665	AG/GA	81/3	106126	C/T	88/4
81680	G/A	81/3	120777	T/A	45/3
81684	T/C	81/3	123761	T/A	81/1
81687	C/T	81/3	123783	A/G	77/3
81699	G/C	64/3	123997	C/T	56/3
81711–81712	TT/CA	64/3	124547	T/C	59/1
81716	G/C	64/3	124566	T/C	59/1
81719	A/C	64/3	134386	G/A	62/5
81741	T/C	63/1	134401	G/A	62/5
81743	C/A	63/1	144551	A/G	72/2
81762	C/T	63/1	151745	C/T	56/3
81768	G/A	62/2	151778	T/G	56/3
82518	A/G	68/1	159185	T/C	35/2
82529	T/A	68/1	159213	T/T	35/2
82531	A/G	68/1	159215	G/A	35/2

showed 90.7% identity. No significant homology was found among the other RS.

Repeat sequences have been observed in all large dsDNA viruses, e.g., baculoviruses, entomopoxviruses, insect iridoviruses, ascoviruses, herpesviruses, and adenoviruses (35). They occur at multiple locations along the genome and may serve as origins of DNA replication (43) and/or as enhancers of transcription (25, 26).

Proteins similar to baculovirus-nudivirus structural envelope proteins. Among the 12 ORFs in the GpSGHV sequence showing similarities to baculovirus and nudivirus genes, five of these, SGHV001, SGHV102, SGHV053, SGHV076, and SGHV005, were found to be putative orthologues of the *p74* (*pif-0*), *pif-1*, *pif-2* and *pif-3 per os* infectivity factors (*pifs*) and

to ODV-e66 genes, respectively. The *pif* genes are structural genes common to all nudiviruses and baculoviruses, whereas *odv-e66* is specific to lepidopteran baculoviruses. They are known or presumed to encode envelope proteins typical of the *odv* phenotype that are defined as important proteins for oral infectivity of baculoviruses released from occlusion bodies (polyhedra or granules) in the midgut lumen. The four PIF proteins were shown to be essential for oral infectivity (41, 44, 52, 58). PIF-0, PIF-1, and PIF-2 are thought to be external proteins with a C-terminal (PIF-0) or an N-terminal (PIF-1 and PIF-2) hydrophobic transmembrane domain assumed to anchor these proteins in the ODV envelope (19, 67), whereas the highly hydrophobic N-terminal sequences of ODV-e66 and PIF-3 would anchor these proteins to the inner side of the

TABLE 2. Potentially expressed ORFs in GpSGHV

ORF	Position	Length (amino acids)	Intergenic distance (bp) ^a	Best BLASTP match							Conserved domain(s) or signature(s) ^c
				Name, source	Accession no.	Score	E value	Identity (%)	Length (amino acids)	MegAlign identity (%) ^b	
SGHV001	1>2091	696	7	p74 protein, <i>Spodoptera litura</i> NPV	CAA09849	90	6.E-16	131/656 (19)	657	20.8	TM, baculo-p74
SGHV002	3071<2088	327	-4								
SGHV003	3785<3234	183	162								
SGHV004	4448<4185	87	399								
SGHV005	5439<4378	353	-71	ODV-E66, <i>Epiphyas postvittana</i> nucleopolyhedrovirus	NP_203211	62	2.E-08	46/192 (23)	680	15.3	TM TM, SP, lyase, Baculo-E66
SGHV006	6576<5491	361	51								DUF676 hydrolase, lipase
SGHV007	6714>7778	354	137								
SGHV008	8618<7803	271	24								Y045_METJA 13-219
SGHV009	8631>10871	746	12	Hypothetical protein MAL7P1.132, <i>Plasmodium falciparum</i> 3D7	XP_001349148	50	6.E-04	60/225 (26)	2,041	22.4	
SGHV010	14205<10891	1,104	19	ORF MSV156, <i>Melanoplus sanguinipes</i> EPV	NP_048227	67	9.E-09	98/449 (21)	1,127	15.4	Coiled-coil region, kinetochore Spc7
SGHV011	14780<14202	192	-4								
SGHV012	15897<14836	353	55								
SGHV013	16552<16376	58	478								
SGHV014	16895<16719	58	166								TM, SP
SGHV015	17011<16853	52	-43								TM
SGHV016	17068>17436	122	56								
SGHV017	18627<17449	392	12								ZFC3HC4 ring
SGHV018	18941>19168	75	313								
SGHV019	21510<20179	443	1,010	Hypothetical protein TA04315, <i>Theileria annulata</i> strain Ankara	XP_954808	55	9.E-06	32/131 (24)	791	18.7	
SGHV020	23524>24054	176	2,013								
SGHV021	25329<24049	426	-6								
SGHV022	25337>26368	343	7								
SGHV023	26422>27552	376	53								AARP2CN (NUC121) domain SM00785
SGHV024	29113<27557	518	4	Mv-ORF74 peptide, <i>Maruca vitrata</i> MNPV	YP_950804	49	1.E-03	31/103 (30)	224	24.6	TM
SGHV025	29205>29957	250	91								
SGHV026	30329>31375	348	371								
SGHV027	31398>32774	458	22	Chitinase Chit1 precursor, <i>Glossina morsitans morsitans</i>	AAL65401	226	3.E-57	129/365 (35)	460	33	TM, Glycoside hydrolase family 18 catalytic domain
SGHV028	33269>33862	197	494								
SGHV029	33890>34573	227	27								Staphylococcal AgrD protein SM00794
SGHV030	34978<34577	133	3								
SGHV031	35917<35060	285	81								HDAC interaction domain, histone deacetylase
SGHV032	36744<35965	259	47								
SGHV033	38089<37043	348	298								
SGHV034	39187<38138	349	48	ORF AMV 260, <i>Amsacta moorei</i> EPV	NP_065042	44	2.E-02	45/173 (26)	504	19.2	
SGHV035	39460<39197	87	9	ORF MSV 238, <i>Melanoplus sanguinipes</i> EPV	NP_048309	57	3.E-07	28/80 (35)	292	33.3	
SGHV036	40088<39741	115	280	ORF 67, shrimp white spot syndrome virus thymidylate synthase	NP_477589	62	1.E-08	28/78 (35)	289	37.4	
SGHV037	40351<40187	54	98								
SGHV038	44374<40850	1,174	498								SP
SGHV039	45446<44439	335	64								TM, SP
SGHV040	45768>48473	901	-1	ORF AMV 130, <i>Amsacta moorei</i> EPV	NP_064912	44	5.E-02	44/141 (31)	1,384	19.1	
SGHV041	49765<48524	413	50	ORF MSV 214, <i>Melanoplus sanguinipes</i> EPV, SCG gene family protein	NP_048285	44	2.E-02	61/319 (21)	386	18.2	

Continued on following page

TABLE 2—Continued

ORF	Position	Length (amino acids)	Intergenic distance (bp) ^a	Best BLASTP match						Conserved domain(s) or signature(s) ^c	
				Name, source	Accession no.	Score	E value	Identity (%)	Length (amino acids)		MegAlign identity (%) ^b
SGHV042	50183<49809	124	43								
SGHV043	50652<50218	144	34								TM
SGHV044	51903<50824	359	171								TM
SGHV045	57046<51860	1,728	-44								SP
SGHV046	57316>58917	533	269								PPASE, inorganic pyrophosphatase signature
SGHV047	58933>60147	404	15	Hypothetical protein CBG22662, <i>Caenorhabditis briggsae</i>	CAE74824	56	4.E-06	57/311 (18)	743	16.6	
SGHV048	60195>60986	263	47	ORF 033, <i>Heliocoverpa armigera</i> nucleopolyhedrovirus G4 ADP- pyrophosphatase	NP_075102	59	4.E-07	56/231 (24)	238	18.8	
SGHV049	60935>61207	90	-52								
SGHV050	61314>62189	291	106								
SGHV051	64184<62202	660	12	ORF MSV 152, <i>Melanoplus sanguinipes</i> EPV putative core protein	NP_048223	47	3.E-03	52/218 (25)	1,306	20	TM
SGHV052	65228<64308	306	123	Hypothetical protein, <i>Plasmodium falciparum</i>	XP_001351434	53	2.E-05	69/251 (27)	540	23.9	
SGHV053	66323<65241	360	12	Per os infectivity factor 2, <i>Gryllus bimaculatus nudivirius</i>	YP_001111333	66	3.E-09	68/274 (24)	378	21.2	Baculo-44
SGHV054	67507<66473	344	149	ORF AMV 054, <i>Amsacta moorei</i> EPV (putative RNA polymerase- associated transcriptional specificity factor)	NP_064836	45	5.E-03	90/379 (23)	822	29.9	
SGHV055	69398<67500	632	-8	ORF AMV253, <i>Amsacta moorei</i> EPV (possible surface protein)	NP_065035	55	1.E-05	100/456 (21)	485	20.6	Internal repeat
SGHV056	69464>70111	215	65								
SGHV057	70161>71117	318	49	Rhoptry protein, <i>Plasmodium yoelli</i> <i>yoelli</i> strain 17XNL	XP_725453	57	2.E-06	51/167 (30)	2,664	30.9	
SGHV058	71829<71593	78	475								
SGHV059	75616<74639	325	2809	RpoD, <i>Plasmodium falciparum</i>	CAA64574	55	5.E-06	74/308 (24)	960	28.9	TM, SP
SGHV060	76249<75623	208	6								
SGHV061	77789<76305	494	55								
SGHV062	77753>90874	4,373	-37	ORF 147, <i>Tricoplusia ni</i> ascovirus 2c	YP_803369	123	3.E-25	271/1374 (19)	1,481	22.5	
SGHV063	92270<90876	464	1								
SGHV064	94234<92447	595	176	ORF AMV130, <i>Amsacta moorei</i> EPV (putative ATP- binding cassette transporter)	NP_064912	44	4.E-02	81/354 (22)	1,384	22.4	TM, Pfam:CBF, SP
SGHV065	98511<94246	1,421	11	ORF AMV039, <i>Amsacta moorei</i> EPV (putative ATPase/DNA helicase)	NP_064821	46	2.E-02	72/302 (23)	532	22.8	
SGHV066	98557>98877	106	45								
SGHV067	99700<98921	259	43								
SGHV068	100046<99720	108	19								
SGHV069	100917<100105	270	58								TM, SP
SGHV070	102415<101105	436	187								
SGHV071	102502>104328	608	86								
SGHV072	105120<104311	269	1,808								TM
SGHV073	105372<105133	79	12								
SGHV074	105419>107557	712	46	ORF 105, <i>Choristoneura occidentalis</i> GV (helicase-2)	YP_654526	45	1.E-02	25/75 (33)	461	20.6	ATPases
SGHV075	108415<107600	271	-1								

Continued on facing page

TABLE 2—Continued

ORF	Position	Length (amino acids)	Intergenic distance (bp) ^a	Best BLASTP match							Conserved domain(s) or signature(s) ^c
				Name, source	Accession no.	Score	E value	Identity (%)	Length (amino acids)	MegAlign identity (%) ^b	
SGHV076	108439>109074	211	23	ORF101, <i>Helicoverpa zea</i> single nucleocapsid NPV per os infectivity factor 3	NP_542724	64	4.E-09	38/154 (24)	199	21.9	TM Pfam:DUF666, SP
SGHV077	109141>112320	1,059	66	ORF AMV, 130 <i>Amsacta moorei</i> EPV (putative ATP-binding cassette transporter)	NP_064912	45	4.E-02	80/368 (21)	1,384	17.8	
SGHV078	113291<112581	236	260								TM
SGHV079	113342>116203	953	50	Alcelaphine herpesvirus 1 DNA polymerase	NP_065512	240	3.E-61	208/740 (28)	1,026	24.9	
SGHV080	116825<116220	201	16								Coiled-coil region
SGHV081	117346<116831	171	5								TM
SGHV082	117839<117360	159	13								
SGHV083	119926<117842	694	2	ORF AMV 214, <i>Amsacta moorei</i> EPV	NP_064996	46	7.E-03	69/271 (25)	404	20.8	
SGHV084	120018>120677	219	2,175								
SGHV085	120898>121665	255	220								
SGHV086	121756>123534	592	90								
SGHV087	123845<123555	96	20								
SGHV088	123967>125925	652	121								C2C2 zinc finger
SGHV089	125940>126482	180	14								
SGHV090	126855>127091	78	372								Coiled-coil region
SGHV091	127188>127994	268	96								TM
SGHV092	128014>128265	83	19								TM, SP
SGHV093	128273>129262	329	7								TM
SGHV094	129284>130105	273	21								
SGHV095	130116>130589	157	10								
SGHV096	130626>131771	381	36								TM, SP
SGHV097	131758>132942	394	-14								TM
SGHV098	132962>133309	115	19								
SGHV099	134004<133522	160	212								
SGHV100	134769<134335	144	330								
SGHV101	134768>135088	106	-2								TM, SP
SGHV102	135141>137099	652	52	Per os infectivity factor 1, <i>Neodiprion abietis</i> nucleopolyhedrovirus	YP_667927	79	1.E-12	92/355 (25)	537	19.4	TM, SP EGF-like domain, TM, SP
SGHV103	138292<137162	376	62								TM, SP
SGHV104	138341>140323	660	48								TM, coiled-coil region
SGHV105	141381<140503	292	179								Coiled-coil region
SGHV106	142793<141384	469	2								Coiled-coil region, Fib-alpha, fibrinogen alpha chain
SGHV107	144369<142804	521	10	Lymphocystis disease virus isolate China cell division protein 48	YP_073712	67	2.E-09	47/156 (30)	690	20.7	
SGHV108	145984<144347	545	-23	Lymphocystis disease virus isolate China cell division protein 48	YP_073712	53	5.E-05	41/162 (25)	690	13.8	
SGHV109	147225<146368	285	383								TM, SP
SGHV110	147867<147262	201	36	mp-nase, <i>Spodoptera litura</i> granulovirus	YP_001256988	60	5.E-08	37/104 (35)	464	28.9	TM, zinc-dependent metalloprotease, matrixin signature peptidase_M10
SGHV111	148623<147964	219	96								TM, SP, zinc protease
SGHV112	149132<148629	167	5								TM, SP
SGHV113	149968<149123	281	-10	Hypothetical protein PY00593, <i>Plasmodium yoelli</i> yoelli strain 17XNL	XP_725532	50	2.E-04	58/231 (25)	1647	23.8	
SGHV114	150262>151584	440	293	ORF MSV016, <i>Melanoplus sanguinipes</i> EPV (leucine-rich repeat gene family protein)	NP_048087	55	7.E-06	88/377 (23)	572	20.5	

Continued on following page

TABLE 2—Continued

ORF	Position	Length (amino acids)	Intergenic distance (bp) ^a	Best BLASTP match						Conserved domain(s) or signature(s) ^c			
				Name, source	Accession no.	Score	E value	Identity (%)	Length (amino acids)		MegAlign identity (%) ^b		
SGHV115	151638>152870	410	53	ORF AMV134, <i>Amsacta moorei</i> EPV (leucine-rich repeat gene family protein)	NP_064916	52	6.E-05	55/236 (23)	535	18.4			
SGHV116	152888>153967	359	17										
SGHV117	154687<153968	239	0	ORF 099L, infectious spleen and kidney necrosis virus	NP_612321	47	1.E-03	19/52 (36)	107	29			
SGHV118	154959<154711	82	23								ZF C3HC4 type, ring finger		
SGHV119	155648<154995	217	35								TM TM, SP ZnF_C2H2 domain, coiled-coil region		
SGHV120	155692>156747	351	43										
SGHV121	157037>157798	253	289										
SGHV122	157992>158342	116	193										
SGHV123	159162>159632	156	819										
SGHV124	160975>162828	617	1,342	ORF 168, <i>Xestia c-nigrum</i> GV	NP_059316	46	6.E-03	20/63 (31)	198	22.2			
SGHV125	163098>163592	164	269	ORF 067, <i>Ecotropis obliqua</i> NPV Cg30	YP_874260	50	6.E-05	35/131 (26)	269	25.9			
SGHV126	163629>163937	102	36	ORF 149, <i>Anticarsia gemmatalis</i> NPV pe38-like	YP_803543	44	3.E-03	25/66 v(37)	209	29.4			
SGHV127	164989>165237	82	1,051								ZnF_C2H2 domain		
SGHV128	165802>166026	74	564								Coiled-coil region		
SGHV129	166271>166507	78	244										
SGHV130	166702>167232	176	194								ORF MSV156, <i>Melanoplus sanguinipes</i> EPV		
SGHV131	167329>169824	831	96		NP_048227	51	4.E-04	114/463 (24)	1,127	19.1			
SGHV132	169957>170157	66	132									EGF-like domain signature 1	
SGHV133	170364>171623	419	206										
SGHV134	171764>172063	99	140									Coiled-coil region, Coiled-coil region, Coiled-coil region, TM	
SGHV135	172083>172313	76	19										
SGHV136	172640>172804	54	326										
SGHV137	172868>173140	90	63										
SGHV138	173153>173515	120	12										
SGHV139	173797>173994	65	281										
SGHV140	174002>175228	408	7										
SGHV141	175407<175141	88	-88										
SGHV142	175631>176704	357	223										
SGHV143	176745>178037	430	40										
SGHV144	179093<178719	124	681	ORF 086, <i>Trichoplusia ni</i> ascovirus 2c	YP_803309	55	1.E-06	25/63 (39)	116	33.6			
SGHV145	179383>180018	211	289	Similar to <i>Plasmodium falciparum</i> trophozoite antigen r45-like protein, <i>Danio rerio</i>	XP_001343112	61	4.E-08	64/174 (36)	334	38.5			
SGHV146	180034>180669	211	15								ORF 179, shrimp white spot syndrome virus		
SGHV147	181201>181731	176	531		NP_477701	45	2.E-03	25/116 (21)	221	19.3			
SGHV148	182102>182929	275	370		NP_477701	55	3.E-06	41/190 (21)	221	19.9			
SGHV149	183239>183601	120	309	Sensory appendage protein 5, <i>Manduca sexta</i>	AAF16716	53	6.E-06	38/113 (33)	231	20.8		Internal repeat	
SGHV150	184602<183613	329	11									Possible surface protein AMV253, <i>Amsacta moorei</i> entomopoxvirus	
SGHV151	184795>185043	82	192										
SGHV152	185143>185514	123	99										
SGHV153	185514>185747	77	-1										
SGHV154	185801>186817	338	53										
SGHV155	186879>187079	66	61										
SGHV156	187447<187274	57	194										
SGHV157	187856<187626	76	178										
SGHV158	188655<187951	234	94										
SGHV159	188579>188788	69	-77										
SGHV160	188838>190025	395	49		NP_065035	50	2.E-04	68/274 (24)	485	21.5			TM

^a -, overlap between adjacent ORFs.^b Amino acid identity based on MegAlign ClustalW analysis of entire ORFs.^c SP, signal peptides; TM, transmembrane domains.

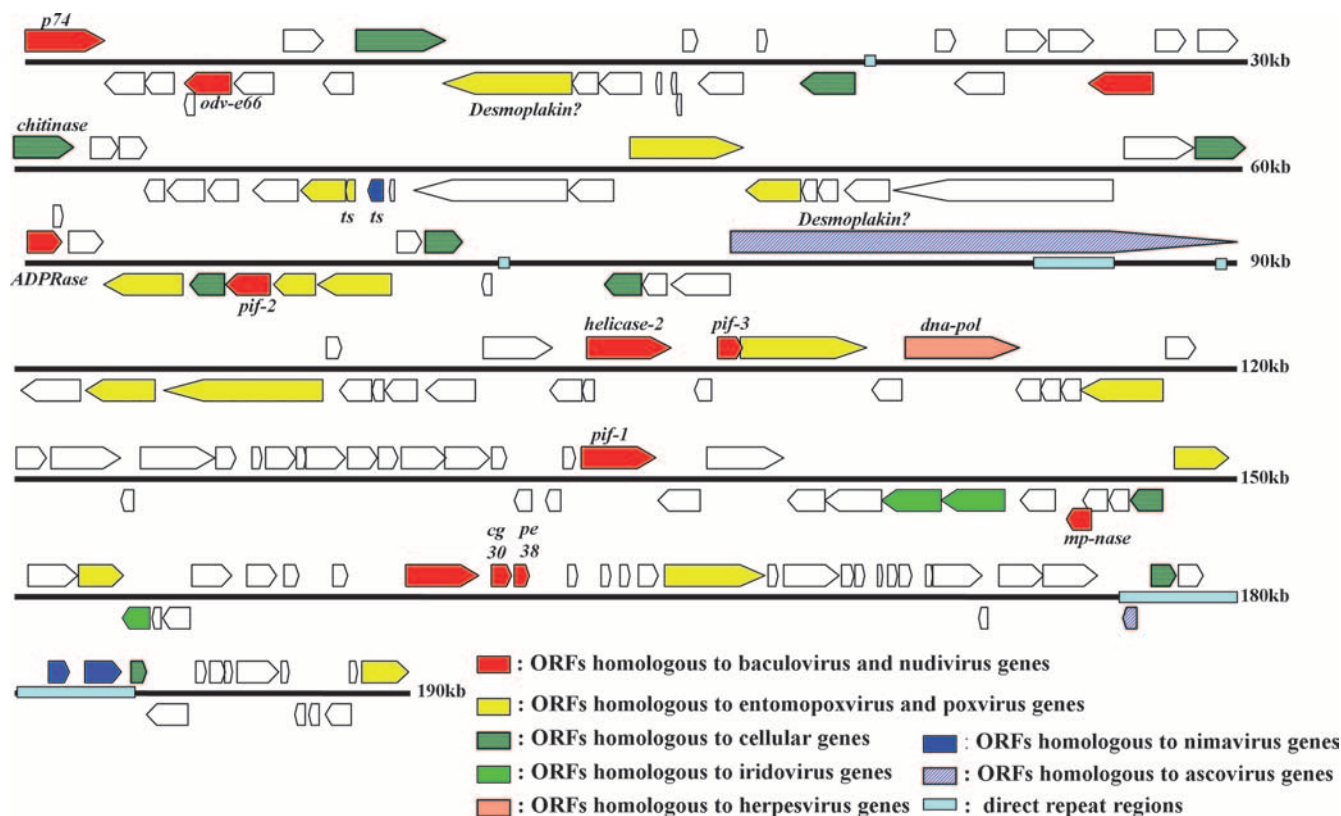


FIG. 3. Linearized representation of the GpSGHV genome. The genome was linearized at the ATG start codon of *p74* (SGHV001). Arrows indicate the positions and directions of transcription for potential ORFs, which are colored according to suggested homology to those of other viruses. Light blue boxes indicate positions of repeat regions.

ODV envelope (11, 32, 66). Interestingly, the SGHV *p74*-like, ODV-e66-like, PIF-1-like, and PIF-3-like proteins display predicted transmembrane domains very similar to those of their baculovirus homologues (Fig. 5). In contrast, the predicted transmembrane domain of SGHV PIF-2-like protein is clearly located in the C-terminal region, whereas it is N terminal in the baculovirus PIF-2 sequence. Whereas PIF-0 and PIF-1 have been identified as attachment proteins that mediate specific binding of ODVs to midgut target cells (44, 52), the role of PIF-2 and PIF-3 in the early events of midgut cell infection is not yet elucidated.

Taken together, these comparisons suggest that the four PIF-like and the ODV-e66-like GpSGHV proteins could be putative orthologues to baculovirus ODV envelope proteins. The *pif-0* to *-3* genes have been identified so far in all sequenced baculoviruses and nudiviruses and are considered core structural genes of these viruses (78). The phylogenetic analyses shown in Fig. 6 did not provide any evidence that any of these *pif* genes was horizontally transferred from a baculovirus or a nudivirus to GpSGHV. Considering the very low sequence similarities, and the consistent phylogenetic trees for the different *pif* genes and assuming that these viruses share a common ancestor, it is suggested that SGHV, nudiviruses, and baculoviruses diverged a long time ago. Nevertheless, the effective presence of these proteins and their localization in the GpSGHV virion first need to be demonstrated before assigning them a putative function.

Putative orthologues to baculovirus-nudivirus auxiliary and unknown proteins.

Baculovirus auxiliary genes are nonessential for replication and virus production but most likely provide a selective advantages to the host-virus relationships under natural conditions (53). Most baculoviruses isolated from lepidoptera have a late-transcribed chitinase that, in association with a cysteine cathepsin, promotes liquefaction of larval tissue and release of occlusion bodies into the environment, thus favoring horizontal transmission (29, 68). We have identified a chitinase gene (SGHV027) showing high BLASTP similarities and significant amino acid identity (33%) to the *Glossina morsitans* midgut chitinase gene. The best BLAST hit to a viral gene was to *Chrysodeixis chalcites* nucleopolyhedrovirus (NPV) chitinase (76), with only 14.7% amino acid identity. Unlike baculovirus chitinases, SGHV027 protein lacks the C-terminal KDEL endoplasmic reticulum retention motif as well as the highly conserved family 18 glycosyl hydrolase motif SGGWT. Furthermore, phylogenetic analysis confirmed that it is distantly related to its baculovirus homologues but is closer to dipteran chitinases (Fig. 7). These data suggest (i) that GpSGHV probably acquired its chitinase from its host rather than from a prokaryote gene as has been suggested for baculovirus chitinases (28) and (ii) that its role is probably different from that of baculovirus chitinases involved, in association with viral cathepsins, in the liquefaction of larval tissues. It is noteworthy that flies dying from GpSGHV infection do not show symptoms of internal lysis. The role played by the predicted GpS-

TABLE 3. Frequencies of 4- or 5-nt motifs in the 5' upstream regions of the SGHV ORFs compared to the complete genome for the baculovirus AcMNPV^a

Virus and motif length	100 nt upstream				200 nt upstream			
	Motif	Occurrence in genome ^b (% of expected occurrence ^c)	Occurrence in upstream regions (% of expected occurrence)	Relative enrichment in upstream regions	Motif	Occurrence in genome (% of expected occurrence)	Occurrence in upstream regions (% of expected occurrence)	Relative enrichment in upstream regions
AcMNPV	TAAG ^d	393 (29)	90 (114)	4	TAAG ^d	393 (29)	137 (87)	3
	TATA	1,314 (66)	172 (149)	2.3	TATA	1,314 (66)	255 (111)	1.7
	ATAA	1,973 (101)	222 (198)	1.9	ATAA	1,973 (101)	363 (162)	1.6
	ATAT	1,616 (81)	170 (147)	1.8	ATAT	1,616 (81)	268 (116)	1.4
	AGTA	671 (49)	70 (89)	1.8	AGTA	671 (49)	109 (69)	1.4
	AAGG	473 (51)	41 (76)	1.5	GATA ^e	867 (63)	137 (87)	1.4
	GATA ^e	867 (63)	74 (94)	1.5	AATA	2,230 (115)	346 (154)	1.3
	CACT ^f	612 (64)	52 (95)	1.5	CAGT ^g	698 (73)	106 (96)	1.3
	AATA	2,230 (115)	186 (166)	1.4	CTTA	393 (28)	59 (37)	1.3
	ATTA	1,957 (98)	163 (141)	1.4	TCAC ^f	669 (70)	99 (90)	1.3
	GTAT	949 (68)	79 (98)	1.4	GCTA	541 (57)	77 (70)	1.2
	CAGT ^g	698 (73)	58 (105)	1.4	CCCC	190 (42)	27 (52)	1.2
	TAAA	2,716 (140)	222 (198)	1.4	TACC	444 (47)	63 (57)	1.2
	AGGG	233 (35)	19 (50)	1.4	TAGT	737 (53)	104 (64)	1.2
	TAGT	737 (53)	58 (72)	1.4	CACT ^f	612 (64)	86 (78)	1.2
	SGHV 4-mer	TAAG	848 (32)	106 (96)	3.0	TAAG	848 (32)	154 (69)
AGTC		506 (53)	36 (88)	1.7	AGTC	506 (53)	67 (82)	1.6
AGGT		536 (52)	38 (87)	1.7	GTAG	676 (66)	88 (101)	1.5
TAGG		458 (44)	32 (73)	1.6	AGGG	245 (58)	30 (84)	1.4
AGTA		1,515 (58)	103 (93)	1.6	AAGT	1,703 (65)	208 (94)	1.4
CAGT		814 (85)	55 (135)	1.6	AGTA	1,515 (58)	183 (83)	1.4
GCGC		178 (123)	12 (195)	1.6	GCAG	418 (106)	50 (149)	1.4
AAGT		1,703 (65)	113 (102)	1.6	AGCC	347 (94)	41 (131)	1.4
TAGT		1,451 (58)	95 (89)	1.5	TAGG	458 (44)	53 (61)	1.4
GTAG		676 (66)	44 (101)	1.5	AAGC	631 (63)	71 (84)	1.3
GTAA		1,989 (76)	128 (115)	1.5	ATAG	1,655 (63)	186 (84)	1.3
CTTA		848 (36)	54 (54)	1.5	TAGT	1,451 (58)	163 (77)	1.3
ATAA		6,696 (101)	420 (149)	1.5	CTTA	848 (36)	95 (48)	1.3
AAGA		2,215 (81)	135 (117)	1.4	AGGT	536 (52)	60 (69)	1.3
AGCC		347 (94)	21 (135)	1.4	CAGT	814 (85)	91 (112)	1.3
AGTT		1,764 (70)	106 (100)	1.4	AAGA	2,215 (81)	243 (105)	1.3
5-mer	ATAAG	338 (35)	70 (172)	4.9	ATAAG	338 (35)	89 (109)	3.1
	TAAGA	297 (31)	46 (113)	3.7	TAAGA	297 (31)	64 (79)	2.5
	GTCAG	80 (57)	11 (187)	3.2	CAGTC	93 (72)	20 (182)	2.5
	GTAAG	128 (34)	16 (100)	3.0	AGGGC	57 (100)	12 (248)	2.5
	AGGGC	57 (100)	7 (289)	2.9	TCCGC	52 (109)	10 (248)	2.3
	TCCGC	52 (109)	6 (297)	2.7	GTAAG	128 (34)	24 (75)	2.2
	CCTTA	83 (26)	9 (67)	2.6	CTGGC	67 (122)	12 (258)	2.1
	CGCGC	28 (143)	3 (362)	2.5	CGCGC	28 (143)	5 (301)	2.1
	GCGCA	79 (148)	8 (354)	2.4	TAGCC	96 (74)	17 (155)	2.1
	TAAGC	99 (28)	10 (67)	2.4	GGGCG	30 (133)	5 (262)	2.0
	TAAGT	369 (40)	37 (95)	2.4	CCTAA	139 (42)	23 (82)	2.0
	CGCCC	30 (164)	3 (388)	2.4	TAAGT	369 (40)	61 (78)	2.0
	TAAGG	83 (22)	8 (50)	2.3	GTCAG	80 (57)	13 (110)	1.9
	AAGGG	94 (60)	9 (136)	2.3	AGTAG	200 (53)	32 (100)	1.9
	AGCGC	42 (79)	4 (177)	2.2	CCCTG	25 (52)	4 (99)	1.9
	AGGTC	42 (30)	4 (68)	2.2	CAGGG	25 (44)	4 (83)	1.9

^a Only the 15 motifs with the highest relative enrichment are shown for each virus. For AcMNPV, sequences that are part of the consensus TATA box [TATA(a/t)A] are underlined. Bold indicates a *P* value of ≤ 0.05 .

^b Both strands, excluding homologous repeats (present for AcMNPV and SHGV).

^c Occurrence of a 4-mer or 5-mer based on random distribution of nucleotides in the complete genome.

^d Part of the AcMNPV late initiator sequence (a/g/t)TAAG.

^e Part of the AcMNPV upstream activating element with sequence (a/t)GATA(a/t).

^f Part of the AcMNPV downstream activating element with sequence (a/t)CACNG.

^g Sequence of the AcMNPV early initiator sequence CAGT.

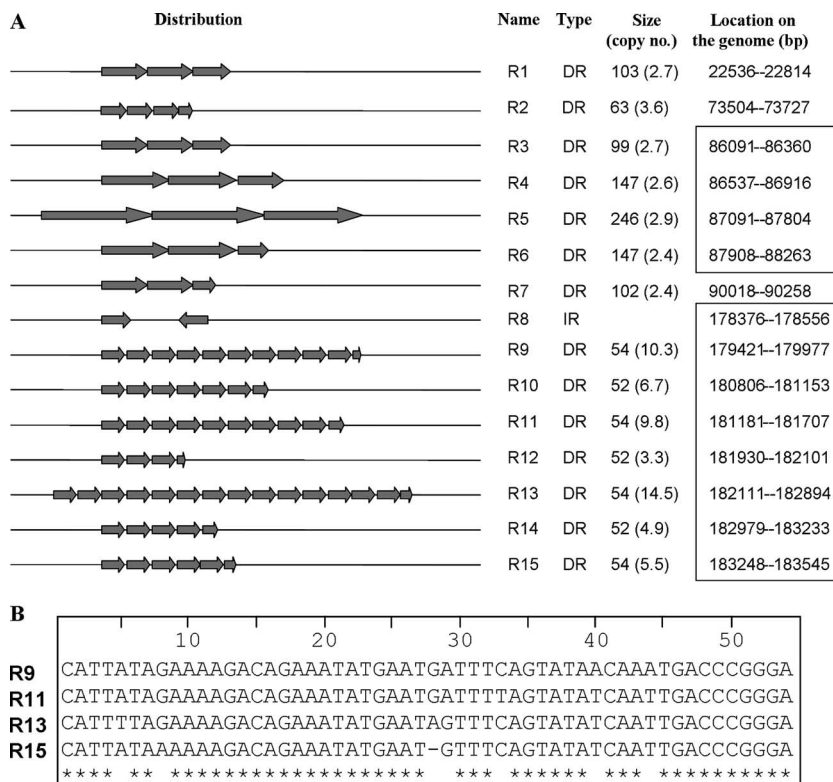


FIG. 4. (A) Loci of the GpSGHV genome with repeated elements. The repeat core elements and their direction are represented by arrows. The name, type, size, copy number, and genome location of the repeat are indicated on the right. DR, direct repeat; IR, inverted repeat. Length, >50 bp; number, >2). (B) Alignment of nucleotide sequences of R9, R11, R13, and R15 direct repeats.

GHV chitinase, assuming that this gene is functional, remains to be elucidated.

The GpSGHV genome has an additional baculovirus-like auxiliary gene, SGHV110, a putative orthologue of *mp-nnase*. This gene, which belongs to the Zn-dependent metalloproteases of the matrix metalloproteinase superfamily, is present in most granulovirus genomes. It is considered to be involved in the breakdown of the peritrophic membrane in the late stage of infection (27, 30, 82). Since no homologues to baculovirus cathepsin or viral enhancing factor genes were detected in the GpSGHV sequence, we hypothesize that the GpSGHV mp-nase-like protein might be involved in the disruption of the peritrophic membrane.

Proteins involved in nucleotide metabolism. Two consecutive genes, the SGHV035 and SGHV036 ORFs, show homology to two putative viral thymidylate synthase genes from *Melanoplus sanguinipes* entomopoxvirus (MSV238, 33.3% identity) and from white spot syndrome virus of shrimp (ORF 67, 37.4% identity), respectively. Thymidylate synthase catalyzes the metabolism of dUTP to yield the nucleotide precursor of dTMP and promotes an important step in the de novo pathway for the biosynthesis of pyrimidines (13). An orthologue of SGHV035 and SGHV036 is also present in *H. zea* nudivirus-1 (14), in *Chilo* iridescent virus (49), and in several herpesviruses (4).

Proteins involved in DNA replication and transcription. The SGHV079 ORF encodes a putative 108-kDa protein homologous to the delta catalytic subunit of DNA polymerases from

several eukaryotic organisms (5). It contains LDFASLYPS, KX3NSXYG, and YGDTS, highly conserved motifs of DNA polymerase of the B family (64). The BLAST best match to a viral gene was to acelaphine herpesvirus DNA polymerase, with 24.9% identity (Table 2), whereas the best hit to a baculovirus had a score of 96 and only 15.9% identity to *Choristoneura fumiferana* MNPV DNA polymerase (76). Alignment and phylogenetic analysis of GpSGHV DNA polymerase with 19 DNA polymerase sequences from different families of large dsDNA viruses provided evidence that GpSGHV is not closely related to any of the previously known families of large dsDNA viruses.

Four GpSGHV ORFs showed weak similarities with four nonconserved baculovirus genes involved in DNA replication. These are SGHV048 with 18.8% identity to ADP pyrophosphatase (ADPRase) of *Helicoverpa armigera* NPV, SGHV074 with 20.6% identity to helicase-2 of *Choristoneura occidentalis* granulovirus, SGHV125 with 25.9% identity to *cg30* of *Ectotropis obliqua* NPV, and SGHV126 with 29.4% identity to *pe38* of *Anticarsia gemmatalis* NPV (Table 2).

An ADPRase gene has been found only in lepidopteran NPVs and granuloviruses and not in hymenopteran and dipteran specific baculoviruses (36). ADPRases belong to a subfamily of Nudix proteins that catalyze the breakdown of ADP-ribose to AMP and ribose-5 phosphate (47). Sequence alignment of SGHV048 with the cellular ADPRases clearly shows that it contains the consensus Nudix motif, G/X5/E/X7/REU/X/EE/X2/U. Deletion of this gene from the *Autographa*

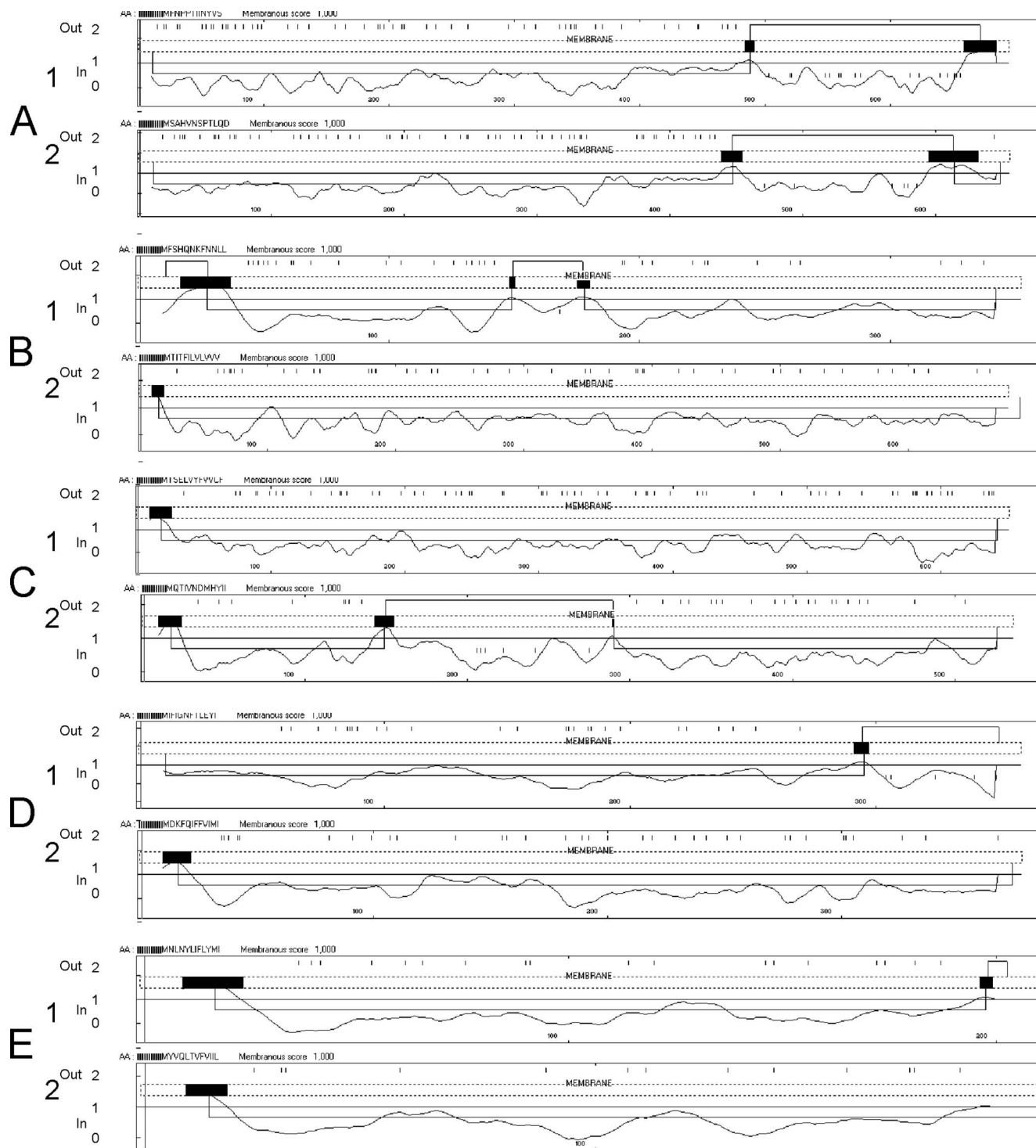


FIG. 5. Hydrophobicity profiles of GpSGHV PIF- and ODV-E66-like proteins compared to their insect virus homologues. (A) SGHV001 (1) and *p74* of *Spodoptera litura* NPV (2); (B) SGHV005 (*odv-e66*) (1) and *odv-e66* of *Epiphyas postvittana* NPV (2); (C) SGHV102 (1) and *pif-1* of *Neodiprion abietis* NPV (2); (D) SGHV053 (1) and *pif-2* of *Gryllus bimaculatus* nudivirus (2); (E) SGHV076 (1) and *pif-3* of *Helicoverpa zea* SNPV (2).

californica MNPV (AcMNPV) sequence resulted in a significant increase in DNA replication in the early stage of infection, but its precise function in virus replication is unclear (23). This type of enzyme is also found in poxviruses (D10 gene of vaccinia virus), where it is considered to be involved in the down-

regulation of viral gene expression (55). The helicase-2 gene (homologous to SGHV052) (data not shown in Table 2) is specific to granulovirus and nudivirus genomes (79). It is distantly related to the conserved helicase core gene of baculoviruses and is probably involved in DNA repair and recombina-

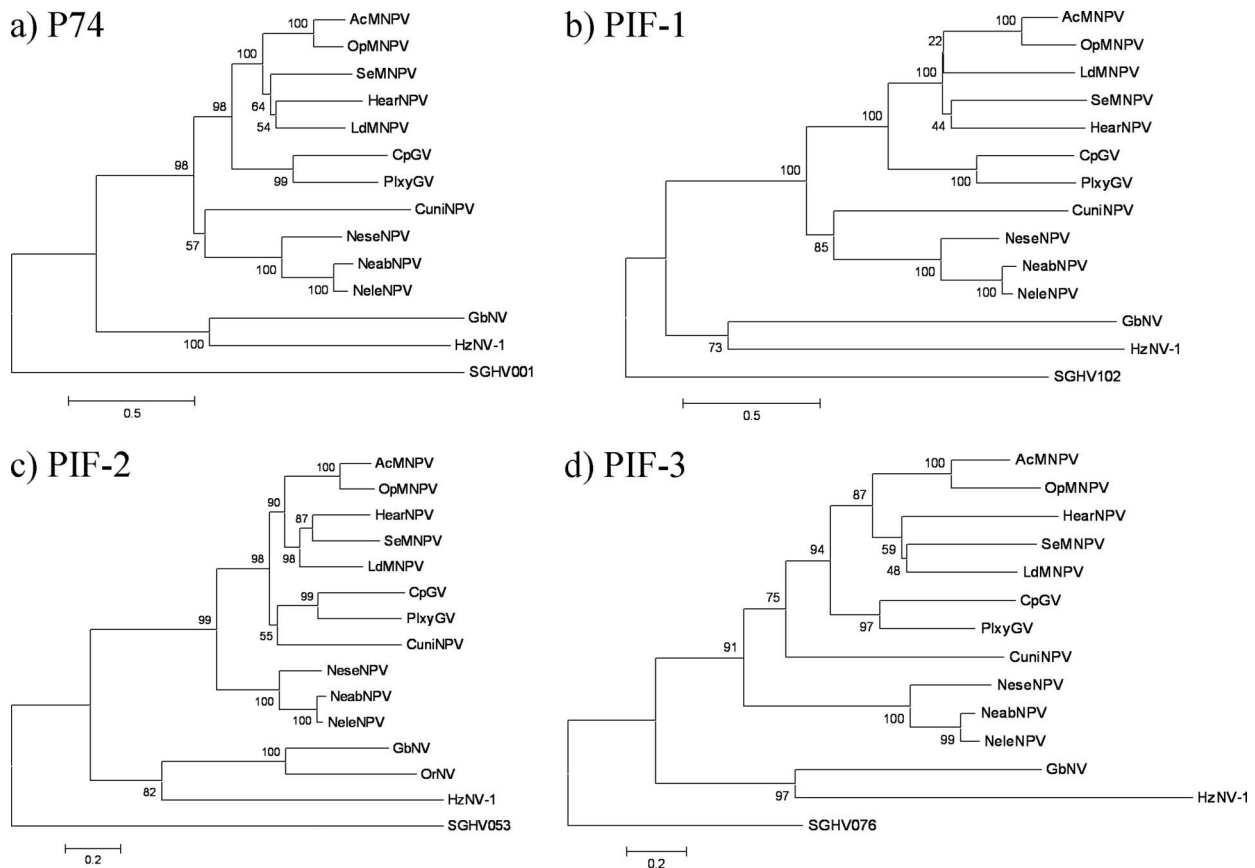


FIG. 6. Neighbor-joining phylogenetic trees of SGHV001 (*p74*) (a), SGHV102 (*pif-1*) (b), SGHV053 (*pif-2*) (c), and SGHV076 (*pif-3*) (d) and their homologues in baculoviruses and nimaviruses. The following viruses (GenBank accession numbers are in parentheses) were included: lepidopteran-specific NPVs AcMNPV (NC_001623), *O. pseudotsugata* (Op) MNPV (NC_001875), *L. dispar* (Ld) MNPV (NC_001973), *S. exigua* (Se) MNPV (NC_002169), and *H. armigera* (Hear) NPV (NC_002654); granuloviruses *C. pomonella* (Cp) GV (NC_002816) and *P. xylostella* (Plxy) GV (NC_002593); hymenopteran-specific NPVs *N. sertifer* (Nese) NPV (NC_005905), *N. lecontei* (Nele) NPV (NC_005906), and *N. abietes* (Neab) NPV (NC_008252); dipteran-specific NPV *C. nigripalpus* (Cuni) NPV (NC_003084); and nudiviruses *H. zea* (Hz) NV-1 (NC_004156), *G. bimaculatus* (Gb) NV (NC_009240), and *O. rhinoceros* (Or) NV (DQ665871 and DQ665870). Distances were calculated using Poisson correction. The homogeneous substitution pattern among lineages with gamma distributed rate among sites (gamma parameter 2.25) was employed for reconstruction of the trees. The robustness of the tree was tested using bootstrap analysis (500 replicates). Numbers on the nodes indicate bootstrap values.

tion (57). *pe38* is an immediate-early baculoviral transactivator gene stimulating early promoter activity and DNA replication (83), whereas *cg-30* has characteristics of eukaryotic transcriptional activator (71). Finally, SGHV065 has 22.8% identity with the AMV039 ORF, encoding a putative ATPase-DNA helicase of *Amsacta moorei* EPV.

Like those of most large DNA viruses, the GpSGHV genome is assumed to encode a DNA-dependent RNA polymerase necessary for the transcription of late genes. This function is normally fulfilled by a complex of proteins including at least two RNA polymerase subunits and initiation, elongation, and termination factors. Surprisingly, none of the 160 putative ORFs of the GpSGHV sequence showed significant hits to DNA-dependent RNA polymerase subunits from large circular or linear dsDNA invertebrate viruses. The only similarities found concerned SGHV059, which had a BLASTP hit to a DNA-dependent RNA polymerase of *Plasmodium falciparum* with 28.9% identity, and SGHV054 showing 29.9% identity to the AMV054 ORF, a putative RNA polymerase-associated transcriptional specificity factor of *Amsacta moorei* EPV (Ta-

ble 2). The apparent lack of a DNA-dependent RNA polymerase orthologue in the GpSGHV genome may reflect differences in the modalities of the expression of its late genes compared to other large DNA viruses.

Similarities of two SGHV ORFs to A-type inclusion poxvirus genes. Sixteen (34%) of the 47 SGHV genes showing similarities to other viral genes had best BLASTP scores to entomopoxvirus genes (Table 2). Two of these, SGHV010 and SGHV131, were homologous to a family of proteins originally described in poxviruses as viral A-type (for acidophilic type) inclusion proteins (ATIs) but later found in many unicellular organisms. This name refers originally to large proteinaceous bodies observed in cells infected with cowpox virus (16). ATIs are made up primarily of intracellular mature virions embedded in a large (160-kDa) late protein, named A-type inclusion protein (56). The gene encoding this protein (*ati*) is highly conserved via a truncated version in many orthopoxviruses that do not make inclusions. It is not essential in tissue culture, and its function as a truncated protein is not known (B. Moss, personal communication). SGHV010 had a low (15.8%) iden-

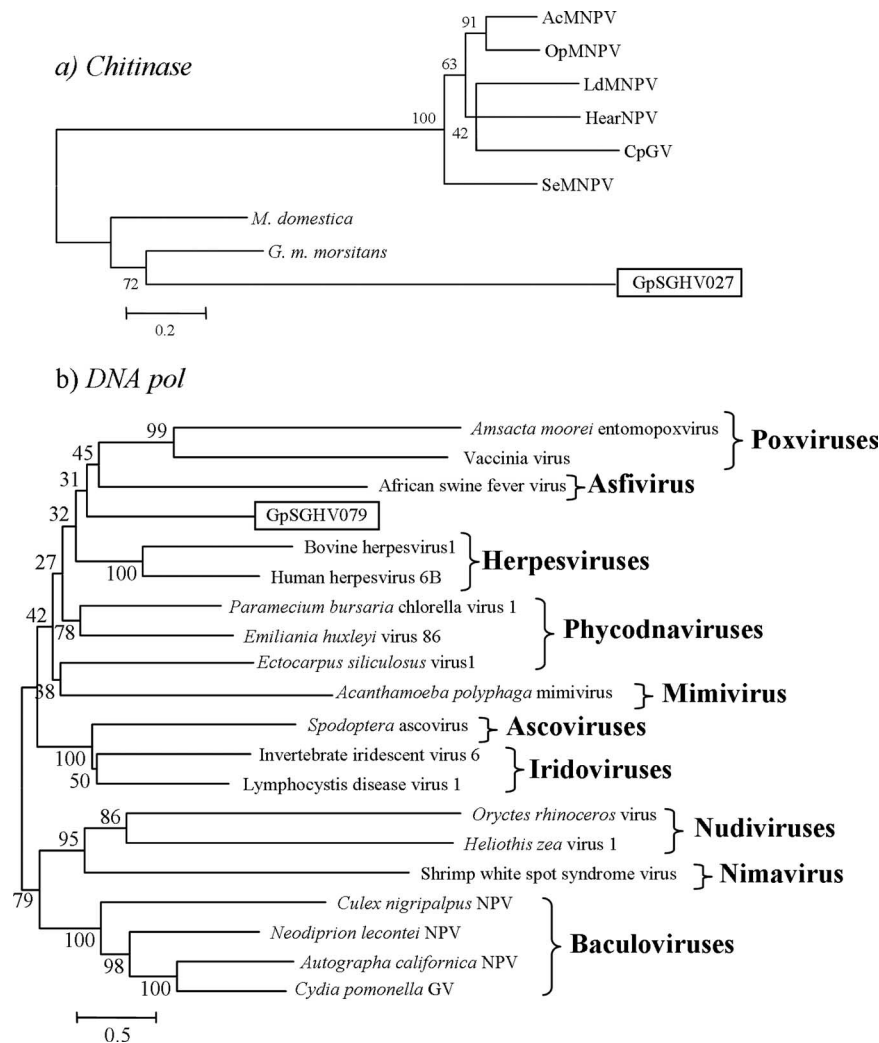


FIG. 7. Neighbor-joining phylogenetic trees of SGHV027 (chitinase) (a) and SGHV079 (DNA polymerase) (b) proteins. The phylogenetic tree of SGHV027 chitinase was constructed by using the following lepidopteran baculovirus chitinases (GenBank accession number are in parentheses): AcMNPV (NC_001623), *O. pseudotsugata* (Op) MNPV (NC_001875), *L. dispar* (Ld) MNPV (NC_001973), *H. armigera* (Hear) NPV (NC_002654), *C. pomonella* (Cp) GV (NC_002816), *S. exigua* (Se) MNPV (NC_002169), and two putative chitinase proteins from *Musca domestica* (ABI29879) and *Glossina morsitans morsitans* (AF337908). Distances were calculated as for Fig. 6. The phylogenetic tree of SGHV079 DNA polymerase and its homologues is based on 2,374 sites of DNA polymerases of 20 viruses from various families. Distances were calculated using Poisson correction. The robustness of the tree was tested using bootstrap analysis (1,000 replicates). Numbers on the nodes indicate bootstrap values. The names of the selected virus families are indicated on the tree. The GenBank accession numbers of the viral DNA polymerases (from top to bottom with the exception of SGHV079) are NP_064832, NP_063712, NP_042783, NP_045328, NP_050219, NP_048532, YP_293784, NP_077578, YP_142676, AAC54632, NP_149500, NP_078724, ABF93350, NP_690550, NP_478036, NP_203396, YP_025217, NP_054095, and NP_148895.

tity with cowpox ATI protein (GenBank accession no. BAA00222). However, alignment of the two proteins revealed a regular distribution of the 194 identical amino acid residues along the homologous region of 1,227 amino acid residues (data not shown). This suggests that the two proteins very likely belong to the same gene family. In addition to poxvirus A-type inclusion proteins, SGHV010 and SGHV131 also shared significant similarities with putative desmoplakins of *Gryllus bimaculatus* nudivirus (ORF77, GenBank accession no. YP_654512) and *Choristoneura occidentalis* granulovirus (ORF 91 GenBank accession no. YP_006242). Desmoplakins are products of cellular genes involved in intracellular junctions, but no function has been assigned to date to their viral homologues. Given these homologies, we suggest that the SGHV010

and SGHV131 proteins could play a role similar to that of poxvirus A-type proteins in embedding virus particles. This could favor cell-to-cell transportation of virions or their protection and dissemination during the extracellular phase of their infection cycle. A similar function could be assigned to desmoplakins of rod-shaped, enveloped, large DNA viruses devoid of an occlusion body (18, 79).

GpSGHV transcripts. We looked for putative GpSGHV mRNAs in the *G. morsitans morsitans* expressed sequence tag (EST) database (<http://www.genedb.org/genedb/glossina/BLAST.jsp>). The tissues used to prepare cDNA came from flies known to be chronically infected with GpSGHV (S. Aksoy, personal communication). The search (TBLASTX) indeed revealed significant identities (64 to 100%) between 12

GpSGHV ORFs and *G. morsitans morsitans* ESTs, mainly and not surprisingly from the salivary gland cDNA library (data not shown). The homologues concerned SGHV010, SGHV044, SGHV045, SGHV060, SGHV061, SGHV062, SGHV064, SGHV074, SGHV107, SGHV108, SGHV111, and SGHV112. Surprisingly, only six of these ORFs (SGHV010, SGHV062, SGHV064, SGHV074, SGHV107, and SGHV108) had hits to viral genes, whereas six (SGHV044, SGHV045, SGHV060, SGHV061, SGHV111, and SGHV112) had no hit with any of the NCBI protein databases. Hence, these 12 ORFs can be considered to encode bona fide functional SGHV proteins.

GpSGHV shares several important features with baculoviruses, nudiviruses, and nimaviruses, such as rod-shaped enveloped virions (Fig. 1), a nuclear site of replication (33), and a large circular dsDNA genome with multiple dispersed repeat regions. However, its genome content and gene arrangement differ in many aspects, and phylogenetic analyses of selected genes did not provide evidence of a close relationship to any of these viruses. In addition, there are many unique biological features of this virus. One important difference concerns the pathology and tissue tropism of GpSGHV. GpSGHV induces a characteristic syndrome of SGH in its host, a property shared with two other viruses of Diptera described in *Musca domestica* (15) and *Merodon equestris* (6). It also induces lesions in gonads and accessory glands of male and female tsetse flies (38, 39, 59, 60) and in the milk glands of females (61), which may be responsible for mother-to-offspring transmission. This specific tissue tropism for some target organs of adult flies reflects a close and very likely ancient adaptation of the virus to the biology of its host. In this respect, the pathology induced by GpSGHV differs profoundly from those described for other large dsDNA viruses of insects.

The analysis of the GpSGHV genome reported here provides a second set of evidence that this virus not only differs from baculoviruses, nudiviruses, and nimaviruses but in fact cannot be assigned to any of the large dsDNA virus families so far known to infect invertebrates or vertebrates. The similarity between the GpSGHV genome and those of other large dsDNA viruses appears to be limited to a very small number of genes with well-known functions, including, surprisingly, four putative orthologues to baculovirus envelope genes *p74*, *pif-1*, *pif-2*, and *pif-3* and a DNA polymerase gene similar to those of herpesviruses. However, the phylogenetic trees of these five GpSGHV genes clearly established their ancient divergence from their putative orthologues. The similarity to about 40 additional viral genes from different virus families and whose function is mostly unknown further underlines the uniqueness of GpSGHV. Taken together these data justify, in our opinion, the proposal that GpSGHV represents a new type of insect virus for which a taxonomic position has to be defined.

ACKNOWLEDGMENTS

We thank Christine Glazer from the Veterinary University, Vienna, Austria, for her help with electron microscopy and Maria Bernery at the Austrian Research Centre for her help with ultracentrifugation. We thank Rudi Boigner for his technical assistance.

REFERENCES

- Abd-Alla, A., H. Bossin, F. Cousserans, A. Parker, M. Bergoin, and A. Robinson. 2007. Development of a non-destructive PCR method for detection of the salivary gland hypertrophy virus (SGHV) in tsetse flies. *J. Virol. Methods* **139**:143–149.
- Afonso, C. L., E. R. Tulman, Z. Lu, E. Oma, G. F. Kutish, and D. L. Rock. 1999. The genome of *Melanoplus sanguinipes* entomopoxvirus. *J. Virol.* **73**: 533–552.
- Aksoy, S., and R. V. M. Rio. 2005. Interactions among multiple genomes: Tsetse, its symbionts and trypanosomes. *Insect Biochem. Mol. Biol.* **35**:691–698.
- Albrecht, J. C., J. Nicholas, D. Biller, K. R. Cameron, B. Biesinger, C. Newman, S. Wittmann, M. A. Craxton, H. Coleman, B. Fleckenstein, and R. W. Honess. 1992. Primary structure of the herpesvirus saimiri genome. *J. Virol.* **66**:5047–5058.
- Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
- Amargier, A., J. P. Lyon, C. Vago, G. Meynadier, and J. C. Veyrunes. 1979. Discovery and purification of a virus in gland hyperplasia of insects. Study of *Merodon equestris* F. (Diptera, Syrphidae). *C. R. Acad. Sci. D* **289**:481–484.
- Asgari, S., J. Davis, D. Wood, P. Wilson, and A. McGrath. 2007. Sequence and organization of the *Heliothis virescens* ascovirus genome. *J. Gen. Virol.* **88**:1120–1132.
- Bawden, A. L., K. J. Glassberg, J. Diggins, R. Shaw, W. Farmerie, and R. W. Moyer. 2000. Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. *Virology* **274**: 120–139.
- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573–580.
- Bideshi, D. K., M. V. Demattei, F. Rouleux-Bonnin, K. Stasiak, Y. Tan, S. Brigot, Y. Bigot, and B. A. Federici. 2006. Genomic sequence of *Spodoptera frugiperda* ascovirus 1a, an enveloped, double-stranded DNA insect virus that manipulates apoptosis for viral reproduction. *J. Virol.* **80**:11791–11805.
- Braunagel, S. C., S. T. Williamson, S. Saksena, Z. Zhong, W. K. Russell, D. H. Russell, and M. D. Summers. 2004. Trafficking of ODV-E66 is mediated via a sorting motif and other viral proteins: facilitated trafficking to the inner nuclear membrane. *Proc. Natl. Acad. Sci. USA* **101**:8372–8377.
- Burt, E. 1945. Hypertrophied salivary glands in *Glossina*: evidence that *G. pallidipes* with this abnormality is particularly suited to trypanosome infection. *Ann. Trop. Med. Parasitol.* **39**:11–13.
- Carreras, C. W., and D. V. Santi. 1995. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**:721–762.
- Cheng, C. H., S. M. Liu, T. Y. Chow, Y. Y. Hsiao, D. P. Wang, J. J. Huang, and H. H. Chen. 2002. Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. *J. Virol.* **76**:9024–9034.
- Coler, R. R., D. G. Boucias, J. H. Frank, J. E. Maruniak, A. Garcia-Canedo, and J. C. Pendland. 1993. Characterization and description of a virus causing salivary gland hyperplasia in the housefly, *Musca domestica*. *Med. Vet. Entomol.* **7**:275–282.
- Downie, A. W. A. 1939. Study of the lesions produced experimentally by cowpox virus. *J. Pathol. Bacteriol.* **48**:361–379.
- Ellis, D. S., and I. Maudlin. 1987. Salivary gland hyperplasia in wild caught tsetse from Zimbabwe. *Entomol. Exp. Appl.* **45**:167–173.
- Escasa, S. R., H. A. Lauzon, A. C. Mathur, P. J. Krell, and B. M. Arif. 2006. Sequence analysis of the *Choristoneura occidentalis* granulovirus genome. *J. Gen. Virol.* **87**:1917–1933.
- Faulkner, P., J. Kuzio, G. V. Williams, and J. A. Wilson. 1997. Analysis of p74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity *in vivo*. *J. Gen. Virol.* **78**:3091–3100.
- Feldmann, H. U., H. F. Barnor, and R. Acs. 1992. Abweichungen in der Reproduktion von *Glossina morsitans submorsitans* Newstead (Diptera: Glossinidae): Untersuchungen zur Behebung eines gestörten Geschlechterverhältnisses und zum Übertragungsweg von Fertilitäts-reduzierenden Viren an die Nachkommen. *Mitt. Dtsch. Ges. Allg. Angew. Ent.* **8**:248–251.
- Feldmann, U. 1994. Guidelines for the rearing of tsetse flies using the membrane feeding technique, p. 449–471. *In* J. P. R. Ochieng-Odero (ed.), *Techniques of insect rearing for the development of integrated pest and vector management strategies*. ICIPE Science Press, Nairobi, Kenya.
- Feldmann, U., V. A. Dyck, R. C. Mattioli, and J. Jannin. 2005. Potential impact of tsetse fly control involving the sterile insect technique, p. 701–723. *In* V. A. Dyck, J. Hendrichs, and A. S. Robinson (ed.), *Sterile insect technique. Principles and practice in area-wide integrated pest management*. Springer, Dordrecht, The Netherlands.
- Ge, J., Z. Wei, Y. Huang, J. Yin, Z. Zhou, and J. Zhong. 2007. AcMNPV ORF38 protein has the activity of ADP-ribose pyrophosphatase and is important for virus replication. *Virology* **361**:204–211.
- Gouteux, J. P. 1987. Prevalence of enlarged salivary glands in *Glossina palpalis*, *G. pallicera*, and *G. nigrofusca* (Diptera: Glossinidae) from the Vavoua area, Ivory Coast. *J. Med. Entomol.* **24**:268.
- Guarino, L. A., M. A. Gonzalez, and M. D. Summers. 1986. Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **60**:224–229.
- Guarino, L. A., and M. D. Summers. 1986. Interspersed homologous DNA

- of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *J. Virol.* **60**:215–223.
27. Hashimoto, Y., T. Hayakawa, Y. Ueno, T. Fujita, Y. Sano, and T. Matsumoto. 2000. Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* **275**:358–372.
 28. Hawtin, R. E., K. Arnold, M. D. Ayres, P. M. Zanotto, S. C. Howard, G. W. Gooday, L. H. Chappell, P. A. Kitts, L. A. King, and R. D. Possee. 1995. Identification and preliminary characterization of a chitinase gene in the *Autographa californica* nuclear polyhedrosis virus genome. *Virology* **212**:673–685.
 29. Hawtin, R. E., T. Zarkowska, K. Arnold, C. J. Thomas, G. W. Gooday, L. A. King, J. A. Kuzio, and R. D. Possee. 1997. Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology* **238**:243–253.
 30. Hayakawa, T., R. Ko, K. Okano, S. I. Seong, C. Goto, and S. Maeda. 1999. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**:277–297.
 31. Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**:215–219.
 32. Hong, T., M. D. Summers, and S. C. Braunagel. 1997. N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. *Proc. Natl. Acad. Sci. USA* **94**:4050–4055.
 33. Jaenson, T. G. T. 1978. Virus-like rods associated with salivary gland hyperplasia in tsetse, *Glossina pallidipes*. *Trans. R. Soc. Trop. Med. Hyg.* **72**:234–238.
 34. Jakob, N. J., K. Müller, U. Bahr, and G. Darai. 2001. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of *Chilo* iridescent virus. *Virology* **286**:182–196.
 35. Jancovich, J. K., J. Mao, V. G. Chinchar, C. Wyatt, S. T. Case, S. Kumar, G. Valente, S. Subramanian, E. W. Davidson, J. P. Collins, and B. L. Jacobs. 2003. Genomic sequence of a ranavirus (family Iridoviridae) associated with salamander mortalities in North America. *Virology* **316**:90–103.
 36. Jehle, J. A., M. Lange, H. Wang, Z. Hu, Y. Wang, and R. Hauschild. 2006. Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **346**:180–193.
 37. Jordan, A. M. 1986. Trypanosomiasis control and African rural development. Longman, London, United Kingdom.
 38. Jura, W. G. Z. O., T. R. Odhiambo, L. H. Otieno, and N. O. Tabu. 1988. Gonadal lesions in virus-infected male and female tsetse, *Glossina pallidipes* (Diptera: Glossinidae). *J. Invertebr. Pathol.* **52**:1–8.
 39. Jura, W. G. Z. O., L. H. Otieno, and M. M. B. Chintawi. 1989. Ultrastructural evidence for trans-ovum transmission of the DNA virus of tsetse, *Glossina pallidipes* (Diptera: Glossinidae). *Curr. Microbiol.* **18**:1–4.
 40. Jura, W. G. Z. O., J. Zdarek, and L. H. Otieno. 1993. A simple method for artificial infection of tsetse, *Glossina morsitans morsitans* larvae with the DNA virus of *G. pallidipes*. *Insect Sci. Appl.* **14**:383–387.
 41. Kikhno, I., S. Gutierrez, L. Croizier, G. Croizier, and M. L. Ferber. 2002. Characterization of *pif*, a gene required for the *per os* infectivity of Spodoptera littoralis nucleopolyhedrovirus. *J. Gen. Virol.* **83**:3013–3022.
 42. Kokwaro, E. D., M. Nyindo, and M. Chintawi. 1990. Ultrastructural changes in salivary glands of tsetse, *Glossina morsitans morsitans*, infected with virus and rickettsia-like organisms. *J. Invertebr. Pathol.* **56**:337–346.
 43. Kool, M., C. H. Ahrens, J. M. Vlak, and G. F. Rohrmann. 1995. Replication of baculovirus DNA. *J. Gen. Virol.* **76**:2103–2118.
 44. Kuzio, J., R. Jaques, and P. Faulkner. 1989. Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. *Virology* **173**:759–763.
 45. Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
 46. Marks, H., X. Y. Ren, H. Sandbrink, M. C. Van Hulst, and J. M. Vlak. 2006. In silico identification of putative promoter motifs of white spot syndrome virus. *BMC Bioinformatics* **7**:309.
 47. McLennan, A. G. 2006. The Nudix hydrolase superfamily. *Cell Mol. Life Sci.* **63**:123–143.
 48. Minter-Goedbloed, E., and D. M. Minter. 1989. Salivary gland hyperplasia and trypanosome infection of *Glossina* in two areas of Kenya. *Trans. R. Soc. Trop. Med. Hyg.* **83**:640–641.
 49. Müller, K., C. A. Tidona, U. Bahr, and G. Darai. 1998. Identification of a thymidylate synthase gene within the genome of *Chilo* iridescent virus. *Virus Genes* **17**:243–258.
 50. Odindo, M. O., P. A. Amutalla, D. A. Turner, E. D. Kokwaro, W. A. Otieno, and D. M. Sabwa. 1982. Morphological variation and incidence of cuticular lesions in the tsetse *Glossina pallidipes* Austen, *G. brevipalpis* Newstead and *G. austeni* Newstead (Diptera: Glossinidae) on the Kenyan coast. *Insect Sci. Appl.* **3**:65–71.
 51. Odindo, M. O., C. C. Payne, N. E. Crook, and P. Jarret. 1986. Properties of a novel DNA virus from the tsetse fly, *Glossina pallidipes*. *J. Gen. Virol.* **67**:527–536.
 52. Ohkawa, T., J. O. Washburn, R. Sitapara, E. Sid, and L. E. Volkman. 2005. Specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to midgut cells of *Heliothis virescens* larvae is mediated by products of *pif* genes *Ac119* and *Ac022* but not by *Ac115*. *J. Virol.* **79**:15258–15264.
 53. O'Reilly, D. R. 1997. Auxiliary genes of baculoviruses, p. 266–300. *In* L. K. Miller (ed.), *The baculoviruses*. Plenum, New York, NY.
 54. Otieno, L. H., E. D. Kokwaro, M. Chintawi, and P. Onyango. 1980. Prevalence of enlarged salivary glands in wild populations of *Glossina pallidipes* in Kenya, with a note on the ultrastructure of the affected organ. *J. Invertebr. Pathol.* **36**:113–118.
 - 54a. Parker, N. J., and A. G. Parker. Simple tools for assembling and searching high-density picolitre pyrophosphate sequence data. *Source Code Biol. Med.*, in press.
 55. Parrish, S., and B. Moss. 2006. Characterization of a vaccinia virus mutant with a deletion of the D10R gene encoding a putative negative regulator of gene expression. *J. Virol.* **80**:553–561.
 56. Patel, D. D., D. J. Pickup, and W. K. Joklik. 1986. Isolation of cowpox virus A-type inclusions and characterization of their major protein component. *Virology* **149**:174–189.
 57. Pearson, M. N., and G. F. Rohrmann. 1998. Characterization of a baculovirus-encoded ATP-dependent DNA ligase. *J. Virol.* **72**:9142–9149.
 58. Pijlman, G. P., A. J. Pruijssers, and J. M. Vlak. 2003. Identification of *pif-2*, a third conserved baculovirus gene required for *per os* infection of insects. *J. Gen. Virol.* **84**:2041–2049.
 59. Sang, R. C., W. G. Z. O. Jura, L. H. Otieno, and R. W. Mwangi. 1998. The effects of a DNA virus infection on the reproductive potential of female tsetse flies, *Glossina morsitans centralis* and *Glossina morsitans morsitans* (Diptera: Glossinidae). *Mem. Inst. Oswaldo Cruz* **93**:861–864.
 60. Sang, R. C., W. G. Z. O. Jura, L. H. Otieno, R. W. Mwangi, and P. Ogaja. 1999. The effects of a tsetse DNA virus infection on the functions of the male accessory reproductive gland in the host fly *Glossina morsitans centralis* (Diptera: Glossinidae). *Curr. Microbiol.* **38**:349–354.
 61. Sang, R. C., W. G. Z. O. Jura, L. H. Otieno, and P. Ogaja. 1996. Ultrastructural changes in the milk gland of tsetse *Glossina morsitans centralis* (Diptera: Glossinidae) female infected by a DNA virus. *J. Invertebr. Pathol.* **68**:253–259.
 62. Sang, R. C., W. G. Z. O. Jura, L. H. Otieno, P. M. Tukei, and R. W. Mwangi. 1997. Effects of tsetse DNA virus infection on the survival of a host fly *Glossina morsitans centralis* (Diptera: Glossinidae). *J. Invertebr. Pathol.* **69**:253–260.
 63. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 64. Shamoo, Y., and T. A. Steitz. 1999. Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* **99**:155–166.
 65. Shaw, M. K., and S. K. Moloo. 1993. Virus-like particles in Rickettsia within the midgut epithelial cells of *Glossina morsitans centralis* and *Glossina brevipalpis*. *J. Invertebr. Pathol.* **61**:162–166.
 66. Slack, J., and B. M. Arif. 2007. The baculoviruses occlusion-derived virus: virion structure and function. *Adv. Virus Res.* **69**:99–165.
 67. Slack, J. M., E. M. Dougherty, and S. D. Lawrence. 2001. A study of the *Autographa californica* multiple nucleopolyhedrovirus ODV envelope protein p74 using a GFP tag. *J. Gen. Virol.* **82**:2279–2287.
 68. Slack, J. M., J. Kuzio, and P. Faulkner. 1995. Characterization of *v-cath*, a cathepsin L-like proteinase expressed by the baculovirus *Autographa californica* multiple nucleopolyhedrovirus. *J. Gen. Virol.* **76**:1091–1098.
 69. Steelman, C. D. 1976. Effects of external and internal arthropod parasites on domestic livestock production. *Annu. Rev. Entomol.* **21**:155–178.
 70. Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
 71. Thiem, S. M., and L. K. Miller. 1989. A baculovirus gene with a novel transcription pattern encodes a polypeptide with a zinc finger and a leucine zipper. *J. Virol.* **63**:4489–4497.
 72. Tidona, C. A., and G. Darai. 2000. Iridovirus homologues of cellular genes—implications for the molecular evolution of large DNA viruses. *Virus Genes* **21**:77–81.
 73. Tsai, C. T., J. W. Ting, M. H. Wu, M. F. Wu, I. C. Guo, and C. Y. Chang. 2005. Complete genome sequence of the grouper iridovirus and comparison of genomic organization with those of other iridoviruses. *J. Virol.* **79**:2010–2023.
 74. Tyler, S. D., and A. Severini. 2006. The complete genome sequence of herpesvirus papio 2 (*Cercopithecine herpesvirus 16*) shows evidence of recom-

- ination events among various progenitor herpesviruses. *J. Virol.* **80**:1214–1221.
75. **van Hulten, M. C., J. Witteveldt, S. Peters, N. Kloosterboer, R. Tarchini, M. Fiers, H. Sandbrink, R. K. Lankhorst, and J. M. Vlak.** 2001. The white spot syndrome virus DNA genome sequence. *Virology* **286**:7–22.
 76. **van Oers, M. M., M. H. C. Abma-Henkens, E. A. Herniou, J. C. de Groot, S. Peters, and J. M. Vlak.** 2005. Genome sequence of *Chrysodeixis chalcites* nucleopolyhedrovirus, a baculovirus with two DNA photolyase genes. *J. Gen. Virol.* **86**:2069–2080.
 77. **Vreysen, M. J. B., K. M. Saleh, M. Y. Ali, A. M. Abdulla, Z.-R. Zhu, K. G. Juma, A. Dyck, A. R. Msangi, P. A. Mkonyi, and H. U. Feldmann.** 2000. *Glossina austeni* (Diptera: Glossinidae) eradicated on the island of Unguja, Zanzibar, using the sterile insect technique. *J. Econ. Entomol.* **93**:123–135.
 78. **Wang, L., J. Xue, C. P. Seaborn, B. M. Arif, and X. W. Cheng.** 2006. Sequence and organization of the *Trichoplusia ni* ascovirus 2c (*Ascoviridae*) genome. *Virology* **354**:167–177.
 79. **Wang, Y., R. G. Kleespies, A. M. Huger, and J. A. Jehle.** 2007. The genome of *Gryllus bimaculatus* nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. *J. Virol.* **81**:5395–5406.
 80. **Whitnall, A. B. M.** 1934. The Trypanosome infections of *Glossina pallidipes* in the Umfolosi Game Reserve, Zululand. *Onderstepoort J. Vet. Sci. Anim. Ind.* **2**:7–21.
 81. **WHO.** 2001. Scientific working group on African trypanosomiasis (sleeping sickness), WHO/TDR Committee Report. World Health Organization, Geneva, Switzerland.
 82. **Wormleaton, S., J. Kuzio, and D. Winstanley.** 2003. The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology* **311**:350–365.
 83. **Wu, X., S. Stewart, and D. A. Theilmann.** 1993. Characterization of an early gene coding for a highly basic 8.9K protein from the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. *J. Gen. Virol.* **74**:1591–1598.
 84. **Yang, F., J. He, X. Lin, Q. Li, D. Pan, X. Zhang, and X. Xu.** 2001. Complete genome sequence of the shrimp white spot bacilliform virus. *J. Virol.* **75**:11811–11820.